

E. M. Buturla
P. E. Cottrell
B. M. Grossman
K. A. Salsburg

Finite-Element Analysis of Semiconductor Devices: The FIELDAY Program

The FIELDAY program simulates semiconductor devices of arbitrary shape in one, two, or three dimensions operating under transient or steady-state conditions. A wide variety of physical effects, important in bipolar and field-effect transistors, can be modeled. The finite-element method transforms the continuum description of mobile carrier transport in a semiconductor device to a simulation model at a discrete number of points. Coupled and decoupled algorithms offer two methods of linearizing the differential equations. Direct techniques are used to solve the resulting matrix equations. Pre- and post-processors enable users to rapidly generate new models and analyze results. Specific examples illustrate the flexibility and accuracy of FIELDAY.

Introduction

The use of Computer-Aided Design (CAD) is generally recognized as a productivity enhancement method, particularly in the electronics industry. Very Large-Scale Integrated (VLSI) circuits require computer simulation at the process, device, circuit, and system levels to accurately predict cost and function prior to fabrication. In this paper, we describe the FInite-ELEment Device Analysis program (FIELDAY), which is used extensively throughout IBM to simulate semiconductor devices.

FIELDAY is a general-purpose computer program which numerically solves the semiconductor transport equations in one, two, or three dimensions for steady-state or transient operating conditions. A number of algorithms for device simulation have been reported in the literature, but most have been designed to simulate a specific device such as an Insulated-Gate Field-Effect Transistor (IGFET), Charge-Coupled Device (CCD), or bipolar transistor [1-8]. Some algorithms do not solve the complete set of governing equations, while others assume steady-state operating conditions; or, because they use a finite-difference technique, have difficulty modeling irregularly shaped structures [9-16]. Recently, finite-element methods have overcome the latter limitations [11, 12, 17-

19]. With few exceptions [16, 20], most algorithms assume that the device can be described by a two-dimensional cross-section. The FIELDAY program does not have these restrictions. FIELDAY can simulate an arbitrary metal-insulator-semiconductor structure and can model a wide variety of physical effects which are important for the accurate simulation of bipolar and field-effect devices. Applications range from the analysis of the effect of short and narrow channels on the threshold of IGFETs to the transient simulation of heavily doped bipolar transistors.

The motivation for this CAD tool is easy to understand when other options for obtaining the same information are considered. There are two obvious alternatives: the best is to fabricate and characterize devices; the other is to use simpler models which are based either on one-dimensional approximations or on extrapolations of data from devices "similar" to those of interest. Both of these approaches have advantages and limitations when compared to FIELDAY.

Clearly, fabrication and testing of actual devices is the best way to discover all of the implications and limitations

Copyright 1981 by International Business Machines Corporation. Copying is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract may be used without further permission in computer-based and other information-service systems. Permission to *republish* other excerpts should be obtained from the Editor.

of any new technology. While this is required of any technology seriously considered for development, it is probably the most expensive and time-consuming option. A single lot of devices fabricated for a new VLSI technology may cost over one million dollars to fabricate and may take six months to complete. Device simulation is a cost-effective method of determining whether a new technology is worth developing. Once that decision has been made, simulation can substitute for many costly matrix experiments that are normally required to optimize a new process and device structure. This is especially important in VLSI technologies because of the statistical nature of device design. Function must be guaranteed for the large number of devices in modern electronic systems. It is impossible to produce enough experimental hardware to test even the most critical combinations of parameters and structures. Large numbers of devices and significant variations in structure from device to device make a statistical design imperative. Another advantage that simulation offers is certainty of the structure and physical parameters of the device; *i.e.*, device design information can be derived before a new fabrication process has been stabilized and the device can be optimized early in the product development cycle. A final advantage is that the internal operation of any device can be easily examined through multidimensional simulation; experimental techniques can do this only approximately and for very few parameters.

Simpler models and the extrapolation of data from existing device structures cannot accurately predict how an entirely new device structure will behave. This can only be determined by a model based on fundamental physical assumptions. The modeling difficulty is compounded by the near-unity dimensional aspect ratios of devices used in VLSI technologies; *e.g.*, a minimum-size n-channel IGFET has a length, width, and depth of similar dimensions. Thus, three-dimensional modeling is required to accurately describe this device. It is also necessary to describe the transient behavior of bipolar devices, mainly because of three-dimensional effects associated with small emitters.

This paper discusses the FIELDAY algorithm and describes the associated interactive pre-processor and post-processor programs which provide a comprehensive device-design package. The following sections describe the physical model and the numerical algorithm utilized, and outline the interactive pre- and post-processing capabilities. Also presented are specific examples of applications of the FIELDAY program, including the analysis of short and narrow IGFETs, VMOS (Vertical Metal Oxide Silicon) transistors, and bipolar transistors.

Physical model

• Semiconductor transport equations

The characteristics of a semiconductor device are modeled by three coupled, nonlinear partial differential equations [21, 22]. These semiconductor transport equations consist of Poisson's equation, Eq. (1), and the equations of electron and hole current continuity, Eqs. (2) and (3):

$$\nabla^2 \psi + \frac{q}{\epsilon} (p - n + N_D - N_A + N_Q) = 0, \quad (1)$$

$$\nabla \cdot \mathbf{J}_n - q \left(\frac{\partial n}{\partial t} + R_n \right) = 0, \text{ and} \quad (2)$$

$$\nabla \cdot \mathbf{J}_p + q \left(\frac{\partial p}{\partial t} + R_p \right) = 0; \text{ where} \quad (3)$$

$$\mathbf{J}_n = q \{ D_n \nabla n - \mu_n n \nabla (\psi + \Delta V_c) \} \text{ and} \quad (4)$$

$$\mathbf{J}_p = -q \{ D_p \nabla p + \mu_p p \nabla (\psi - \Delta V_v) \}. \quad (5)$$

The three unknown quantities are the space-charge potential (ψ) and the electron (n) and hole (p) mobile charge densities at each instant of time. N_D and N_A are the donor and acceptor impurity densities, N_Q is the density of fixed charged particles, the constant q is the magnitude of electronic charge, ϵ is the dielectric permittivity, and $\mathbf{J}_n$ and $\mathbf{J}_p$ are the electron and hole current densities. R_n and R_p are the electron and hole recombination rate densities, μ_n and μ_p are the electron and hole mobilities, ΔV_c and ΔV_v are changes in the conduction and valence band edges [23] of heavily doped semiconductors, and D_n and D_p are the electron and hole diffusion coefficients.

Poisson's equation relates the space-charge potential to mobile and fixed charges, with mobile charge densities given by Boltzmann statistics. The flow of mobile charge carriers is described by the equations of electron and hole current continuity. The electron and hole mobilities that appear in these equations are functions of electric field strength $|\nabla \psi|$ and impurity density [24]. The diffusion coefficients D_n and D_p are related to the electron and hole mobilities by the Einstein relationship.

In the FIELDAY model, the recombination-generation mechanisms include carrier generation due to avalanche multiplication [25], photo-generation, and Auger and Shockley-Read-Hall recombination [26-28]. These recombination-generation mechanisms couple the two current continuity equations and introduce strong nonlinearities, particularly for the case of avalanche multiplication.

• Boundary conditions

The three semiconductor transport equations with three unknowns require three boundary conditions. Boundary

conditions are specified at contacts and along the entire edge of the semiconductor device model. At contacts, the space-charge potential and the electron and hole densities are required. At ohmic contacts, thermal equilibrium and space-charge neutrality determine charge-carrier densities. Carrier concentrations at Schottky contacts are either set to fixed values [29] or modulated by a thermionic recombination velocity [30].

At the noncontact boundaries of the semiconductor, the normal components of the electron and hole current densities and the electric-field strength are all equal to zero. In insulator regions, only the space-charge potential and its normal derivative are considered.

• Bipolar, unipolar, and no current flow

Many devices can be accurately simulated without modeling the current flow of one or both mobile carriers. Hole current can be ignored in the simulation of an n-channel IGFET under most operating conditions. In this case, the equation for hole current continuity is not used. If the flow of both carrier types can be ignored, as in the simulation of the capacitance of a reverse-biased p-n junction, only Poisson's equation is solved.

The FIELDAY model may operate in one of three modes. The first assumes bipolar current flow, and three unknown quantities (ψ , n , and p) must be determined. The second assumes unipolar current flow with two unknowns (ψ and n), and the third assumes no current flow and uses only one unknown (ψ). In insulator regions, there are no mobile charge carriers and only Poisson's equation is needed.

Numerical approach

The solution of the governing nonlinear, coupled partial differential equations by classic techniques, *i.e.*, integration and application of boundary conditions, is impossible for any except the most basic problems. Instead, an approximation technique is necessary to transform the continuum problem to a discrete one. The unknown variables are determined at a large but finite number of points in both space and time so that an accurate solution of the equations is obtained. In FIELDAY, the finite-element method transforms Poisson's equation, and a hybrid finite-difference finite-element technique transforms the current continuity equations [14]. Euler's method approximates the rate of change of mobile charge density with time. These equations are then linearized by one of two methods. The first decouples the three discrete equations and solves them iteratively [31]. The second, more involved, approach solves the equations simultaneously using Newton's method [12, 32]. Either

approach results in large, sparse matrix equations which must be solved numerous times to obtain the final solution.

• Poisson's equation

The finite-element method transforms Poisson's equation from a continuous to a discrete form. Computationally convenient piecewise approximations over arbitrary regions or elements are constructed. Using appropriate energy conservation principles, an elemental expression is obtained which relates the unknown ψ to element properties and charge density.

The following functional is used to approximate Poisson's equation [33]:

$$I = \int_V \frac{\epsilon}{2} (\nabla\psi)^2 dv - \int_V q\psi(p - n + N_D - N_A + N_Q)dv, \quad (6)$$

where V is the volume of the domain. The functional represents the energy of the system and may be expressed as the sum of a large but finite number of energies for all of the elements. Equation (6) is then applied to a single element over which a linear variation of ψ with respect to position is assumed. At equilibrium the energy is a minimum, and the first variation with respect to ψ is zero. By taking the first variation and integrating over space, the following expression is obtained:

$$[A]\{\psi\} = [B]\{p - n + N_D - N_A + N_Q\}. \quad (7)$$

The matrix $[A]$ is symmetric and its terms are a function of the element geometry and permittivity. The number of nodes in an element is the order of $[A]$. For two-dimensional triangular, and for three-dimensional tetrahedral and right-prismatic elements, the order is 3, 4, and 6, respectively. The vectors $\{\psi\}$ and $\{p - n + N_D - N_A + N_Q\}$ represent the space-charge potential and charged-particle densities at each node of the finite element.

The matrix $[B]$ is the element volume distribution matrix, which relates the portion of area or volume to a particular node of an element. The form of the distribution matrix is extremely important. The traditional procedure, common in structural analysis and usually denoted as the lumped method, assumes equal portions of area or volume associated with each node. The consistent method assumes the same approximation for charge density as for the potential, in this case a linear variation [34]. Either scheme usually results in anomalous oscillations of potential with respect to position. In the semiconductor problem, this effect is compounded by the exponential rela-

tionship between mobile charge density and potential [35]. To avoid these anomalies, FIELDAY uses a unique approach. Within an element, the area or volume closest to a node is associated with that node. For a two-dimensional triangular element, the area is defined by the perpendicular bisectors of the sides of the element. For a three-dimensional prismatic element, the volume associated with each node is the appropriate portion of the triangular face times one-half the altitude of the element. For a three-dimensional tetrahedral element, the volume associated with each node is described by the intersections of the planes that are perpendicular bisectors of the edges of the element. It should be noted that use of tetrahedra should be limited since they produce asymmetric results; they should be used only where prismatic elements cannot adequately describe the geometry of the problem.

The elemental matrix, Eq. (7), is applied to each element of the domain and all contributions are combined into a global matrix. The order of the global matrix is the total number of nodes approximating the domain. Boundary conditions are then applied and solution of the modified global matrix equation will then yield the value of ψ at each nodal point.

• Current continuity equations

The current continuity equations are solved using a hybrid scheme. The finite element division of space is used along with a difference approximation to describe current flow between nodes. The electron continuity equation is transformed using Gauss' theorem:

$$\int_V \nabla \cdot \mathbf{J}_n dv = \int_S \mathbf{J}_n \cdot d\mathbf{S} = \int_V q \left(\frac{\partial n}{\partial t} + R_n \right) dv. \quad (8)$$

Here, the surface S encloses the volume V . Constant current density is assumed within each element. The current flowing between two nodes is the product of the current density and its flux cross-section. The flux cross-sections are lines or areas defined by the perpendicular bisectors of the edges of the elements. Note that this scheme is consistent with the discrete approximation of Poisson's equation. The electron particle current flowing from node i to node j along side k is

$$I_k = q\mu_n \frac{kT}{q} \frac{d_k}{l_k} \{n_j Z(\Delta_k) - n_i Z(-\Delta_k)\}, \quad (9)$$

where d_k and l_k are the flux cross-section and length of side k , and μ_n is the electron mobility for the element. The Bernoulli function is defined by

$$Z(\Delta_k) = \frac{\Delta_k}{e^{\Delta_k} - 1}, \quad (10)$$

where Δ_k is the potential difference along the k th side. Evaluation of the Bernoulli function about the point $\Delta_k = 0$ requires special attention.

With the use of Euler's method to approximate the rate of change of mobile carrier density with respect to time, the discrete elemental electron current continuity equation becomes

$$\left([\mathbf{C}] + [\mathbf{B}] \frac{1}{\Delta t} \right) \{n(t + \Delta t)\} = [\mathbf{B}] \left\{ \frac{n(t)}{\Delta t} - R_n \right\}. \quad (11)$$

Note that $[\mathbf{C}]$ and R_n are evaluated at time t . The matrix $[\mathbf{C}]$ is nonsymmetric and its coefficients are of the form $\mu(d/l)Z(\Delta)$.

The global form of the electron current continuity matrix equation is assembled on an element-by-element basis. The total number of unknowns may be less than the Poisson matrix equation, since current flows only in the semiconductor. A similar development is followed for the hole current continuity equation. The following is the discrete representation of the governing differential equations:

$$\{F_1\} = [\mathbf{A}]\{\psi\} - [\mathbf{B}]\{p - n + N_D - N_A + N_Q\}, \quad (12)$$

$$\{F_2\} = \left([\mathbf{C}] + [\mathbf{B}] \frac{1}{\Delta t} \right) \{n(t + \Delta t)\} - [\mathbf{B}] \left\{ \frac{n(t)}{\Delta t} - R_n \right\}, \text{ and} \quad (13)$$

$$\{F_3\} = \left([\mathbf{D}] + [\mathbf{B}] \frac{1}{\Delta t} \right) \{p(t + \Delta t)\} - [\mathbf{B}] \left\{ \frac{p(t)}{\Delta t} - R_p \right\}. \quad (14)$$

• Linearization scheme

Since the discrete equations are nonlinear and coupled, a linearization scheme is required to solve them. Two algorithms have been implemented in FIELDAY. Both require an initial guess of the solution followed by an adjustment of the guess according to certain criteria until Eqs. (12)–(14) are satisfied to an acceptable degree of accuracy. The first method, described by Gummel, decouples the three equations and solves them serially. The second technique uses Newton's method to linearize the three equations and solves them simultaneously. Each approach has its own merits.

The decoupled approach, which solves F_1 , F_2 , and F_3 serially [31], is attractive since portions of the program can be written and tested independently. The disadvantage of the approach is possible slow convergence, since for many applications the equations are strongly coupled.

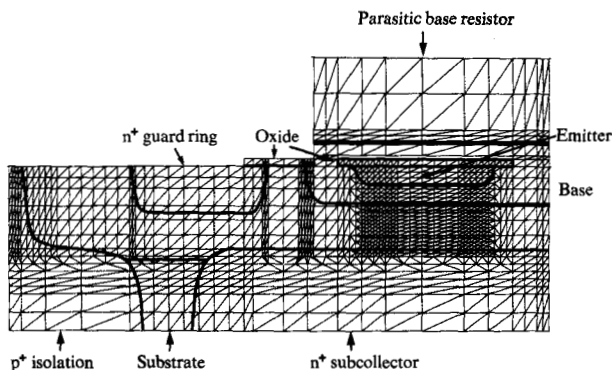


Figure 1 POINTS-generated finite-element mesh used to simulate a bipolar transistor. The parasitic base resistor models the flow of base current around the end of the emitter in the third dimension, which is not simulated.

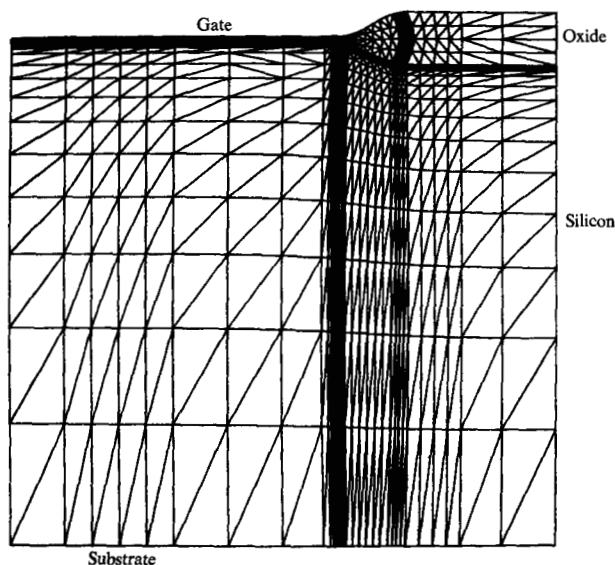


Figure 2 TRIM-generated finite-element model of width cross-section of an IGFET. This model is used to simulate the narrow channel in devices with recessed-oxide field regions.

The coupled method, which solves F_1 , F_2 , and F_3 simultaneously [12, 36], first expands the discrete equations in Taylor-series expansions in terms of ψ , n , and p . The higher-order terms are neglected and an elemental equation is obtained which relates the elemental function derivatives in the Jacobian matrix to the increments of unknowns and the homogeneous equations F_1 , F_2 , and F_3 . Evaluation of the Jacobian matrix requires computing derivatives of a large number of complex terms. As a result, program development is more difficult and time-consuming.

● Matrix solution method

With either linearization scheme, matrix equations must be solved. The matrices are moderately large and their order is usually between 100 and 4000. They are also relatively sparse since the number of nonzero terms is usually less than five percent. The solution of the matrix equations accounts for most of the computational effort, and so a judicious choice of technique is very important.

A direct-matrix-solution technique was chosen rather than an iterative technique since the solving time for iterative methods strongly depends on the numerical conditioning of the matrix and may fail to converge in some situations. Direct methods require much more storage, but the solution time for different problems of the same order will not vary drastically and will always yield a set of results. The approach in FIELDAY is to use a symbolic and numeric factorization procedure for direct solution of the appropriate matrix equations [37].

The decoupled approach requires the solution of three matrix equations of the order N by N . The coupled approach requires the solution of a single $3N$ -by- $3N$ equation. Since the computational work is proportional to the square of the order of the matrix, the coupled approach requires more computational effort per iteration. However, the decoupled approach may require more iterations. The question of which approach should be used has been addressed in a previous work, where it was shown that the decoupled approach was more efficient for "weakly" nonlinear problems and the coupled approach more efficient for "highly" nonlinear and transient problems [36].

Three-dimensional problems require much more computational resource than two-dimensional problems. For a mesh of the same degree of accuracy, the three-dimensional solution requires of order N^3 more CPU time and storage where N is the number of planes in the three-dimensional model. Since N may typically be equal to or greater than 10, the three-dimensional analyses are significantly more expensive.

● Modified Newton's method

A simple modification of the Newton technique can dramatically reduce the amount of time spent solving the matrix equations. In the modified method, the Jacobian matrix is computed and factored only every m th iteration. The factor m is determined internally within FIELDAY and is adjusted as the solution proceeds. Significant computational savings result. In a series of six test problems, total computation time was cut in half. In fact, the savings are greatest for larger problems since a greater part of their execution time is spent solving the

matrix equations. There are additional savings that result when using the modified Newton approach for a sequence of similar problems. This occurs for transient or steady-state analyses with similar boundary conditions. In these cases, the Jacobian from the preceding problem can be reused, which again saves computational effort.

Pre- and post-processing

Significant time and cost benefits are achieved by using interactive graphics. A package of interactive programs has been designed to execute on the IBM 3277 Display Station Graphics Attachment. These programs are used to generate complex finite-element models and to analyze results. For pre-processing, the model-generation programs fall into three categories: model definition, impurity concentration designation, and input verification.

Model definition consists of generating a finite-element mesh, assigning material properties to each element, and designating contact nodes. The FIELDAY user has several options available for model definition. POINTS is a semiautomatic generation scheme. A rectangular grid work is defined with uniform spacings, which are changed as required. The model is synthesized from rectangular regions, each with a constant permittivity, and then scaled to the problem dimensions. The user also supplies information indicating the element material type and contact position. A bipolar mesh generated with POINTS is shown in Fig. 1. To create this mesh by hand takes about one month; to generate it with POINTS and its interactive graphics capability requires no more than a few hours of the user's time. POINTS generates, displays, and stores the mesh for later use by FIELDAY. During mesh display, the contact nodes are indicated, and interactive windowing and node and element numbering may be utilized.

Another option for model definition is TRIM (TRIangular Mesh generator). Here, the user specifies the boundary of a region and selects a mathematically regular grid, such as a rectangular mesh. TRIM generates a conformal map of that mesh onto the user's model. The same mesh may be mapped onto geometrically similar models; thus the user need not respecify the mesh-generation information but only a few details relative to his model. Contact designation, mesh storage, and display features are similar to those available for POINTS. Figure 2 shows a mesh, with an unusual semi-recessed oxide shape, generated by TRIM. Note the varying density of elements over the model. This increases the accuracy of solution at points where the fields are rapidly changing.

A library of FIELDAY models exists for frequently modeled devices with well-defined structure. With these

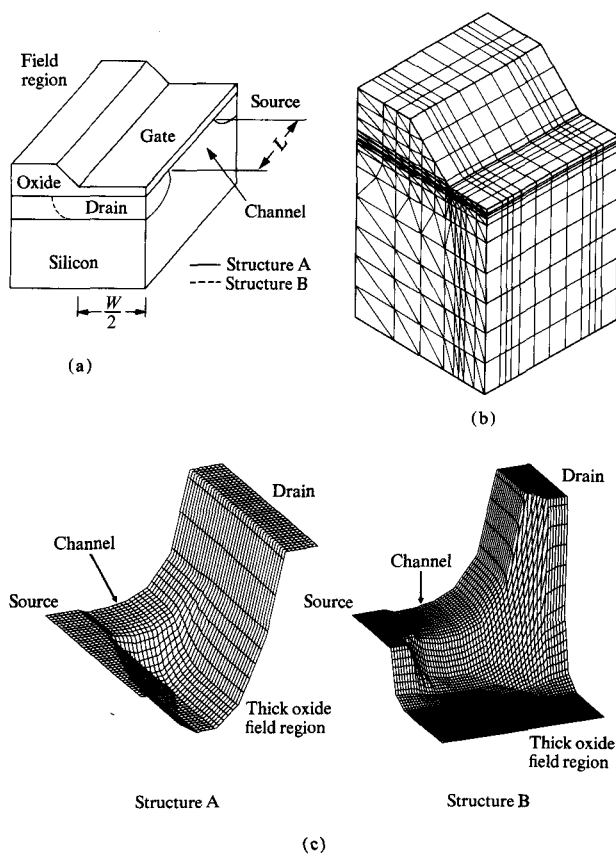


Figure 3 (a) Structure of a short and narrow IGFET device. Oxide thicknesses are 50 and 800 nm for the active and field devices, respectively. The source and drain are abrupt, cylindrical junctions 0.5 μm deep. The channel doping may be described by $C = C_0 \exp [-(Y - R)^2/2S^2] + C_b$, where $C_0 = 2.5 \times 10^{16} \text{ cm}^{-3}$, $S = 0.2 \mu\text{m}$, $C_b = 1.0 \times 10^{15} \text{ cm}^{-3}$, and $R = 0.0 \mu\text{m}$.

(b) The finite-element mesh used to simulate a 1.5-by-1.5- μm short and narrow IGFET containing 13 planes of 164 nodes each.

(c) Modeled surface potential for Structures A and B at a drain-to-source bias of 5.0 V and a source-to-substrate bias of 1.0 V. The device length and width is 1.5 μm .

models, the user supplies various parameters such as oxide thickness and junction depth, and the model is stretched to reflect the given parameters. Again, mesh storage and display are possible. For three-dimensional simulations, a two-dimensional model is created using one of the previously described techniques and is replicated in the third dimension to produce a mesh of right-triangular prisms. The user can then delete elements, change their material properties, or re-assign contacts. Figure 3 shows a short and narrow IGFET and the mesh used for simulation.

Impurity concentration is designated by assigning an electrically active impurity ion density to each node of the

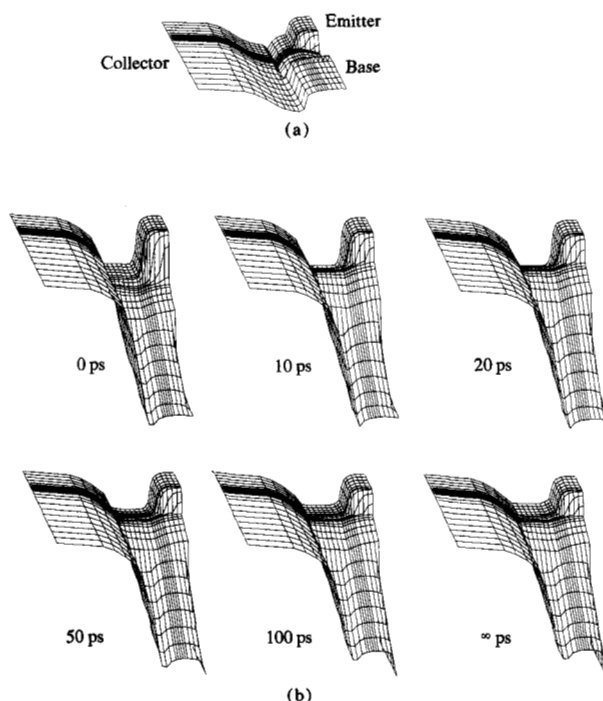


Figure 4 (a) Perspective plot of the $\log |N_A - N_D|$ for a bipolar transistor. (b) Perspective plots of $\log(n)$ following a base voltage step. The gradual increase of electron density under the emitter shows that base "pinch" resistance is limiting the device's switching speed.

finite-element mesh. This is accomplished by specifying measured values, employing a process simulator, or describing a profile as the sum of analytic expressions. The pre-processing program, DOPING, allows viewing of the impurity concentration profile prior to FIELDAY execution to ensure that the device being modeled is the desired one. The impurity concentration can be displayed with contour plots, line graphs, or perspective plots. Figure 4(a) shows a perspective plot of the doping profile for a bipolar transistor.

Input verification consists of mesh checking and input consistency checking. During mesh checking, the finite-element mesh is examined for errors and poorly shaped elements. The areas of the mesh in which problems occur are highlighted, and the user can interactively window to determine how to modify the mesh. The types of errors that can arise include overlapping elements and dangling nodes. Poorly shaped elements are obtuse triangles or elements with large aspect ratios. Other input is examined for completeness and consistency.

For post-processing, a program called FEMPLOT permits rapid interpretation of the FIELDAY analysis with

the ability to interactively view the results. FEMPLOT will display nodal values of potential, electron density, or hole density with contour plots, line graphs, or perspective plots. The elemental values of electric field and current density may also be displayed. Contour plots are a means of displaying nodal data values. Lines of equal value are drawn through points of equal value interpolated along the sides of the elements. Perspective graphs are a means of displaying all the nodal data from a two-dimensional surface of a model. Figure 4(b) shows a series of perspective plots of the log of the electron concentration at various times during the transient response. The gradual increase of electron density under the emitter shows that base "pinch" resistance is limiting the device's switching speed.

Model definition using interactive graphics replaces the time-consuming task of meticulously defining every node and element in the finite-element mesh, which formerly took 65-70% of the total analysis time. FEMPLOT minimizes the time the FIELDAY user spends searching through results on a node-by-node or element-by-element basis. Device designers are able to optimize designs by rapidly viewing simulation results and noting activity within the device that cannot be measured experimentally. Therefore, the use of interactive graphics results in higher engineering productivity by reducing development time, lowering development costs, and making more resources available for product optimization.

Device simulation results

In this section, several applications of the FIELDAY program are described which illustrate its features, accuracy, and flexibility. First, a model of the effect of short and narrow channels on IGFET thresholds is presented. Separate two-dimensional models simulate the short-channel effect for wide devices and the narrow-channel effect for long devices. The accuracy of these simulations is demonstrated by the close agreement between the model and experimental data. A three-dimensional model of an IGFET is also presented. Comparisons of two- and three-dimensional results show the need for this approach for short and narrow devices. A VMOS transistor was optimized using the program. In this case simulation showed that a superior device could be designed. Simulation of the transient response of a bipolar transistor is described, and again excellent agreement between data and model is demonstrated. Three-dimensional transient capability of the program is illustrated by simulation of reverse recovery of an ellipsoidal junction.

• Short- and narrow-channel effects in IGFETs

The FIELDAY program has been frequently used to model short- and narrow-channel effects in IGFETs

[20, 32, 38-41]. The short-channel effect reduces the threshold of an n-channel IGFET. As the drain approaches the source, its positive bias raises the surface potential near the source and thus increases the amount of current flow in the device at a given gate voltage. Narrow channels have the opposite effect. The threshold increases as the edges of the high-threshold field region move toward the center of the lower-threshold active device. Since channel length and width vary from device to device across a VLSI chip, and to a larger extent from chip to chip, a statistical analysis of the effect of these parameters on threshold is required. Variation in channel length may contribute more than 50% to the total threshold tolerance of a well-designed IGFET.

These competing effects are modeled independently for long and narrow, or for short and wide, devices by simulation of mid-channel two-dimensional cross-sections. Short and narrow devices must be simulated by a three-dimensional model. Here, applications of the model to devices as short as $0.7\text{ }\mu\text{m}$ and as narrow as $1.5\text{ }\mu\text{m}$ are discussed.

The effect of short channels was modeled and measured on wide devices designed for a process with minimum feature sizes of $1.0\text{ }\mu\text{m}$. The oxide thickness and doping profile of a capacitor with the same structure as the FET device were measured by a pulsed capacitance technique. This capacitor was simulated with a one-dimensional transient model. Figure 5 shows the agreement between measured capacitance values for this device and those derived from the displacement current density predicted by simulation. The simulated capacitance has been adjusted by a constant value of gate bias. This shift is the sum of the voltage equivalent of the charge found in the insulator and the work function difference between the actual gate material and that assumed in the model. The excellent agreement between the model and the data confirms the accuracy of the model and reinforces the validity of the transient capacitance measurement technique.

Devices with the above structure were simulated for channel lengths of 10.0, 2.0, 1.3, 1.0, and $0.7\text{ }\mu\text{m}$ with a two-dimensional model having 1620 nodes. Also modeled was the inversion charge of the capacitor as a function of gate voltage. Figure 6 shows the modeled and actual threshold as a function of source-to-substrate bias for a $10.0\text{-}\mu\text{m}$ -long device. The effect of decreasing channel length on threshold is illustrated in Fig. 7.

The narrow-channel effect was investigated for the SAMOS (Silicon and Aluminum Metal Oxide Semiconductor) transistor [39, 41, 42]. The threshold of this de-

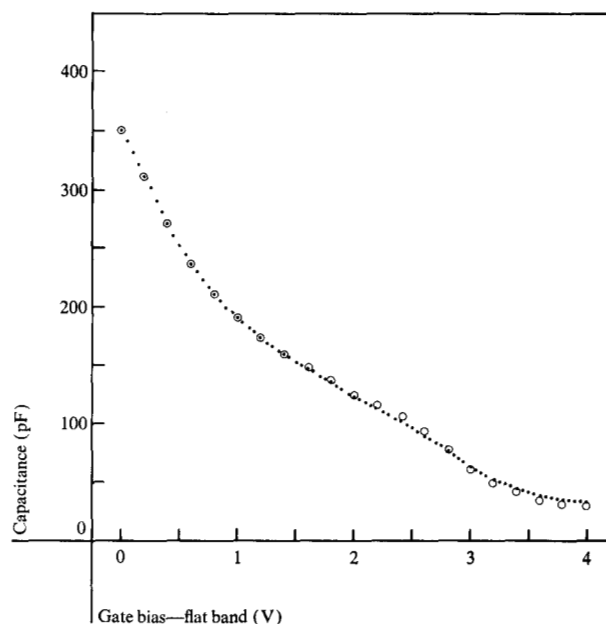


Figure 5 Measured ($\circ$) and simulated ($\bullet$) capacitance versus gate-to-substrate bias. The simulated results have been shifted by -0.56 V to account for work function and oxide charge differences. The channel doping parameters are $C_0 = 5.24 \times 10^{16}\text{ cm}^{-3}$, $C_b = 1.2 \times 10^{15}\text{ cm}^{-3}$, $S = 15.3\text{ nm}$, and $R = -24.0\text{ nm}$. The oxide thickness is 27.8 nm .

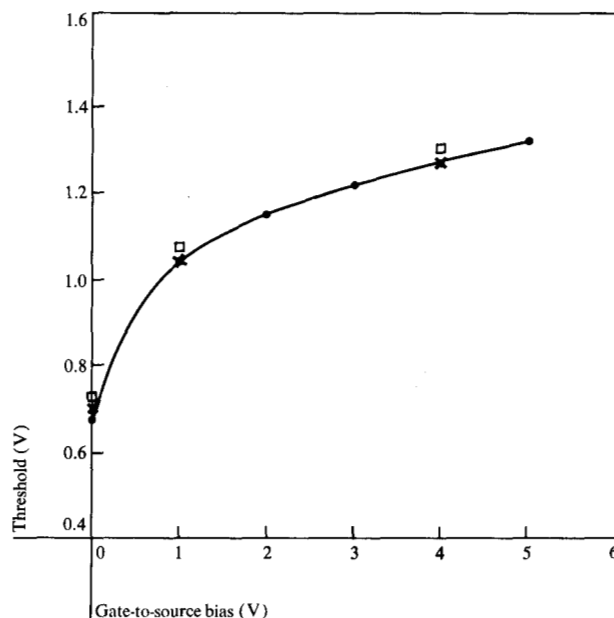


Figure 6 Measured ($\square$) and modeled ($\bullet$ capacitor model, $\times$ IGFET model) long-channel threshold. The simulated threshold has been adjusted by -0.56 V to account for work function and effective oxide charge difference between the actual devices and the model. The threshold of the $10.0\text{-}\mu\text{m}$ device is defined at 40 nA of normalized source current. The threshold of the IGFET as modeled by the capacitor is defined at 10^{10} electrons of inversion charge per square cm. The vertical and lateral junction depths are 0.25 and $0.15\text{ }\mu\text{m}$, respectively.

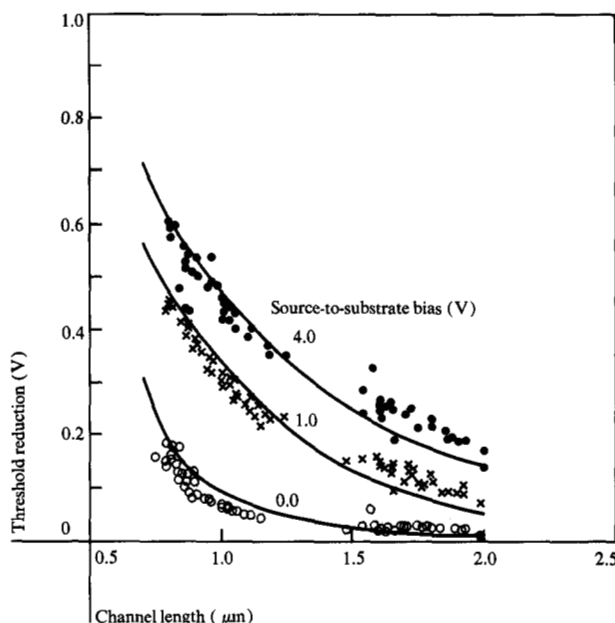


Figure 7 Measured (●, ×, ○) and modeled (—) reduction in threshold with channel length at source-to-substrate biases of 0.0, 1.0, and 4.0 V and a drain bias of 4.0 V. The simulated threshold has been adjusted by -0.56 V to account for work function and effective oxide charge difference between the actual devices and the model.

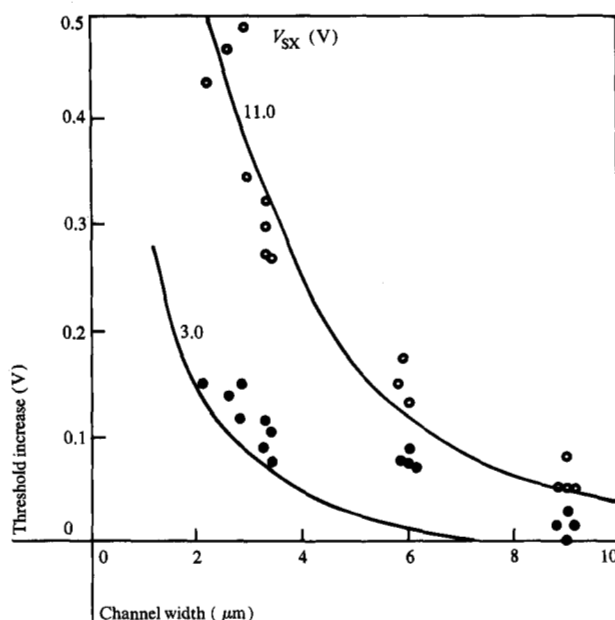


Figure 8 Measured (○, ●) and modeled (—) increase in threshold with decreasing device width for a long-channel SAMOS transistor.

vice was determined by modeling the width cross-section of the transistor. The threshold was calculated by linearly extrapolating the variation of inversion with gate bias to

zero charge. Figure 8 shows the agreement between the empirical and modeled threshold at two substrate biases for long devices.

A three-dimensional simulation [20] of the threshold of a short and narrow device was made on two devices with structures defined in Fig. 3. The difference between the two structures is the shape of the diffused source and drain. In Structure A these diffusions extend under the field oxide, while in Structure B they are terminated at the edge of the thin-oxide region. Figure 3(b) shows the 2132-node finite-element mesh used to model these devices. Simulated subthreshold characteristics were used to define the threshold of a device $1.5 \mu\text{m}$ long and wide. The sensitivity of this threshold to increasing source-to-substrate bias is shown in Fig. 9. Also shown is the threshold as predicted by a composite of separate, two-dimensional, short- and narrow-channel models. For the composite model, the threshold is defined as the algebraic sum of the threshold changes predicted by the independent short- and narrow-channel models and the threshold of a long and wide device. In this case, the width and length cross-sections were taken at mid-channel. The composite model is inadequate because it fails to predict the effect of relatively minor differences between Structures A and B. In addition, the result of the composite model produces a threshold dependence on source-to-substrate bias that bears little functional resemblance to the actual characteristics.

The different behavior of Structures A and B can be easily explained. The extension of the diffusions raises the surface potential in the field-oxide region near the active device. This reduces the impact of the narrow-channel effect and results in a lower threshold for the device with Structure A. Figure 3(c) shows a perspective plot of the surface potential for Structures A and B. Here, the influence of the extensions of the source and drain under the field oxide can be readily seen.

Separate analysis of the short- and narrow-channel effects of a device over a range of sizes and operating conditions takes several weeks of work and approximately 20 CPU hours on an IBM System/370 Model 168. This expense results in about 800 values of current as a function of bias and device size. These can be reduced to 80 values of threshold. It is well worth the cost and effort, as it could take up to six months to obtain similar results empirically, at over 100 times the cost.

• Simulation of VMOS field-effect transistors

VMOS transistors are being investigated for use in one-device random access memories [43, 44]. A VMOS memory cell uses devices which have the shape of inverted

pyramids to charge and discharge the buried-diffusion storage capacitor. A typical cell structure is shown in Fig. 10. The threshold of this transistor must be high enough to prevent discharge of the capacitor when the surface diffusion is grounded and low enough to adequately charge the capacitor when the surface diffusion and the gate are positively biased.

Meeting these criteria with a VMOS device is difficult because of its asymmetric structure. The p-type region above the buried n^+ diffusion raises the threshold of the device when charging the capacitor. This reduces the stored charge. The doping level in this p-type region cannot be reduced because the decreased capacitance of the buried diffusion would also reduce the stored charge. The presence of this layer has an additional effect. It causes the threshold of the transistor to decrease rapidly with increasing voltage on the buried diffusion. This requires an additional increase in the nominal threshold so that the capacitor will not discharge, when the surface diffusion is grounded, and further degrade the stored charge. These effects can be described by a single figure of merit defined as the sum of a holding and a charging loss. As shown in Fig. 11(a), the charging loss is the increase in threshold with increasing surface-diffusion-to-storage-capacitor bias. The holding loss, shown in Fig. 11(b), is the decrease in threshold with increasing storage-capacitor-to-surface-diffusion bias. This loss represents the inefficiency of the transistor caused by dependence of the threshold on source and drain bias.

Various transistor designs were investigated using the FIELDAY program [45]. It was found that the transistor characteristics were very sensitive to the variation of the doping level and position of the p-type region. A very steep profile, along with an additional implanted p-type region under the gate oxide, produced a nearly optimum device. The loss-versus-signal characteristics of this implanted VMOS device, a conventional VMOS transistor, and a planar device with similar oxide thickness and channel length are shown in Fig. 12. With a 5-V power supply, the conventional VMOS device can store only 75% of the charge of a planar device, while the implanted device can store 90%. Thus, a 20% increase in efficiency was predicted through simulation. This may be translated directly into increased memory performance or reduced chip area and cost.

• Transient simulation of bipolar transistors

In most cases the performance and function of IGFET devices can be predicted from their steady-state behavior. The intrinsic speed of these devices exceeds the speed at which practical IGFET integrated circuits can operate. This is because IGFETs are majority-carrier

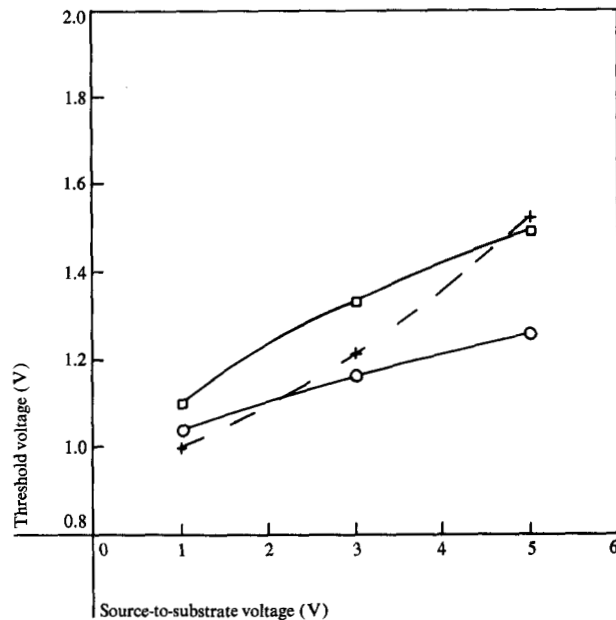


Figure 9 Three-dimensional ($\circ$, $\square$) and composite (+) model threshold versus source-to-substrate bias for a device $1.5 \mu\text{m}$ long and wide with Structures A ($\circ$) and B ($\square$). The device structure is described in Fig. 3. The drain-to-source bias is 5.0 V.

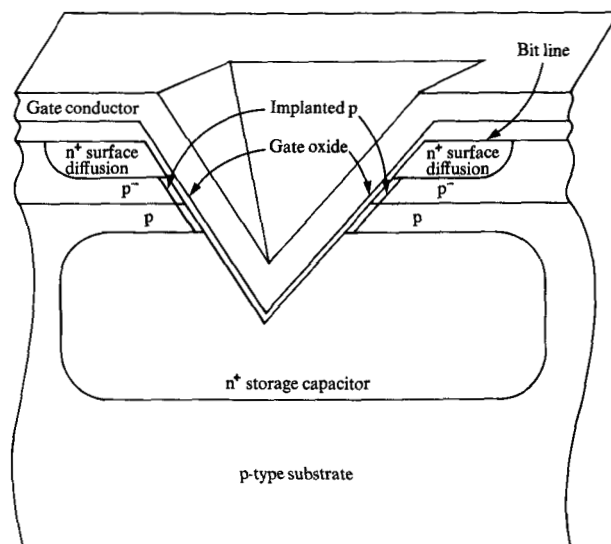
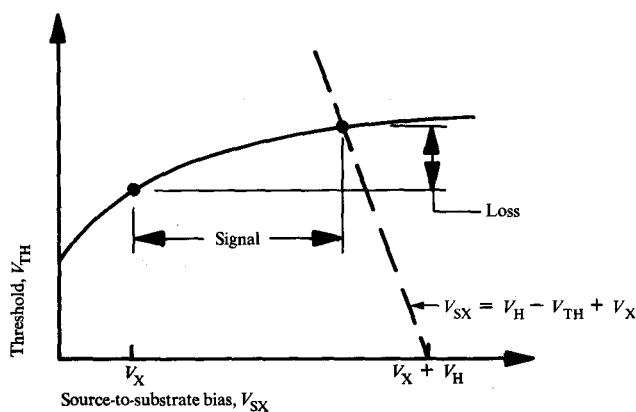
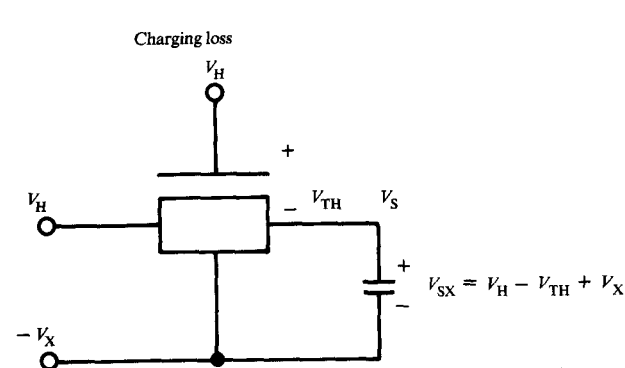
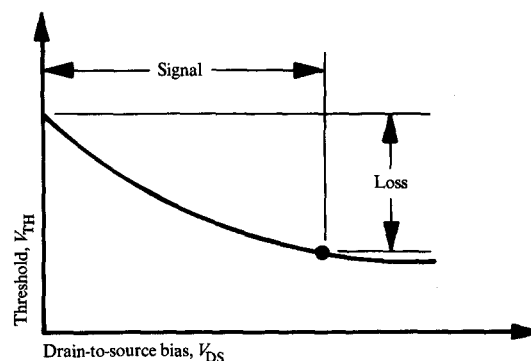
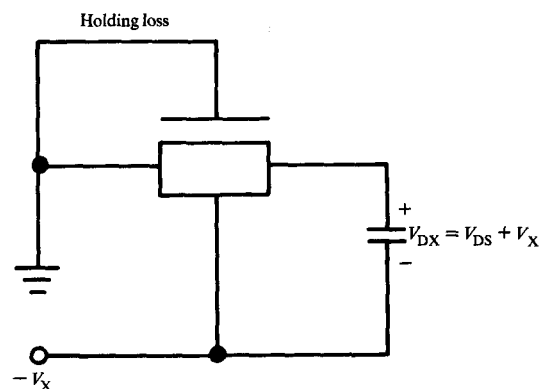


Figure 10 VMOS dynamic memory cell structure.

devices with no significant minority-carrier injection. On the other hand, bipolar devices are minority-carrier devices and their performance in an integrated circuit depends on the transient response of individual devices. Thus, transient simulation of bipolar devices is most important.



(a)



(b)

Figure 11 Schematics of (a) charging loss and (b) holding loss of a one-device memory cell. The charging loss is caused by the increase of threshold with source-to-substrate bias. The holding loss is caused by the reduction in threshold with drain-to-source bias.

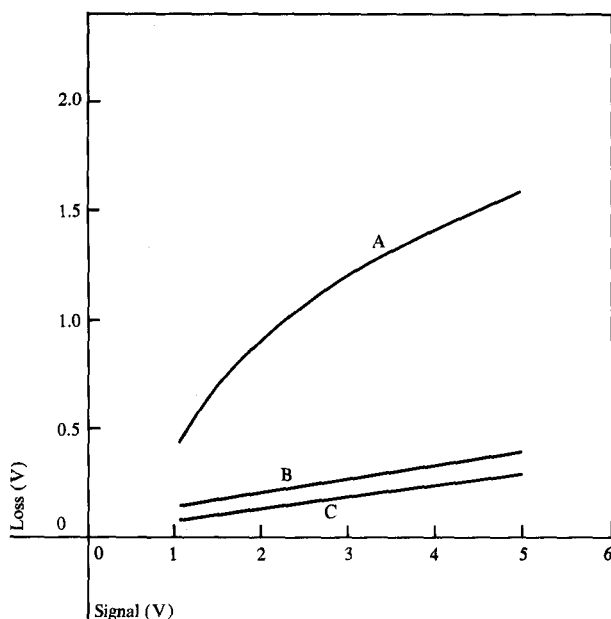


Figure 12 Total holding plus charging loss versus stored voltage for a conventional VMOS transistor (Curve A), an implanted VMOS transistor (Curve B), and an implanted planar transistor (Curve C).

An npn bipolar transistor was simulated in two dimensions with the FIELDAY program [46]. The response of the collector current to a rapidly increasing base voltage was measured and modeled. Figure 13 shows the predicted and actual transient response of the collector voltage. Good correlation is shown between experimental and modeled characteristics.

The transient three-dimensional capability of FIELDAY is demonstrated with the simulation of the reverse recovery of an ellipsoidal junction. A structure similar to that shown in Fig. 14 may be found at the four corners of every integrated bipolar transistor. The reverse recovery of this structure will play an important role in the performance of bipolar devices as the area of emitters is reduced. The results are shown in Fig. 14. To first order, the recovery time is in agreement with that predicted by classical theory:

$$T = W_b^2 / 2D_n = 0.45 \text{ ns}.$$

Although detailed transient analysis of three-dimensional structures is presently quite costly, this example

illustrates the feasibility of such a capability and serves to guide further development of algorithms which should allow routine use.

Summary and conclusions

The capabilities and methods of the FIELDAY program have been described in this work. Several specific examples of one-, two-, and three-dimensional steady-state and transient applications have been presented to illustrate the flexibility of the program. Close correlation of simulation results and experimental data illustrate the accuracy of the model and the credibility of its underlying assumptions and computational methods. Pre-processors speed the creation of new models through interactive mesh generation. Post-processing programs allow rapid examination of the internal operation of devices and subsequent improvement of design.

These capabilities form a comprehensive device CAD tool which allows prediction of the characteristics of new devices and rapid response to problems affecting device function and reliability. In its predictive role, FIELDAY can be used to evaluate new device concepts and to optimize device design prior to fabrication. This is extremely important because of the time and cost involved in evaluating new ideas for integrated devices. FIELDAY, and similar programs, fill a gap in the development of integrated circuits. This gap is between the generation and design of new device concepts and the simulation of circuits using those devices. This role is important for any CAD tool because significant changes and improvements in design can often be made only at the early stages of development. In addition, these tools can replace costly matrix experiments and allow device design and process stabilization to occur simultaneously. Careful simulation can avoid disastrous and costly mistakes that often plague new product development.

In a responsive role, FIELDAY can offer rapid and definitive analysis of device phenomena that limit circuit function or affect the reliability of a product. In this case, a hypothesis can be proposed and tested without fabricating devices, thus reducing by an order of magnitude the time required to solve this type of problem. Simulation allows examination of the internal operation of devices, and the resulting insights often spark innovative solutions.

Acknowledgments

Sincere appreciation is expressed for the efforts of a number of colleagues at the IBM facilities in Essex Junction, Vermont; East Fishkill, New York; and Yorktown Heights, New York. The authors would like to especially thank R. Crook, who encouraged the initial

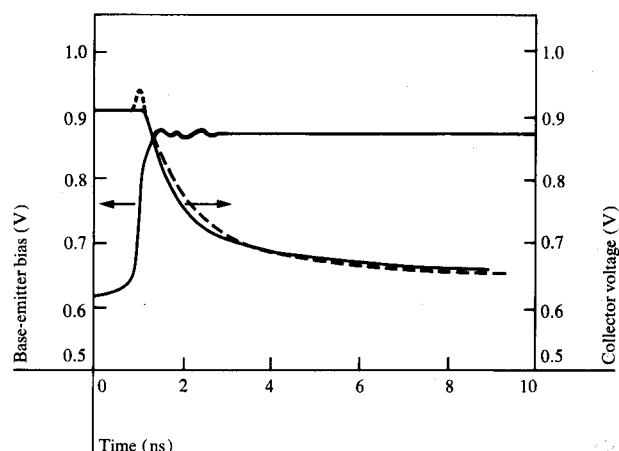


Figure 13 Simulated (—) and measured (---) transient response of a bipolar transistor to an abrupt change in base-emitter bias. The base voltage in the simulation has been decreased by 0.028 V to make the modeled and simulated steady-state collector currents equivalent.

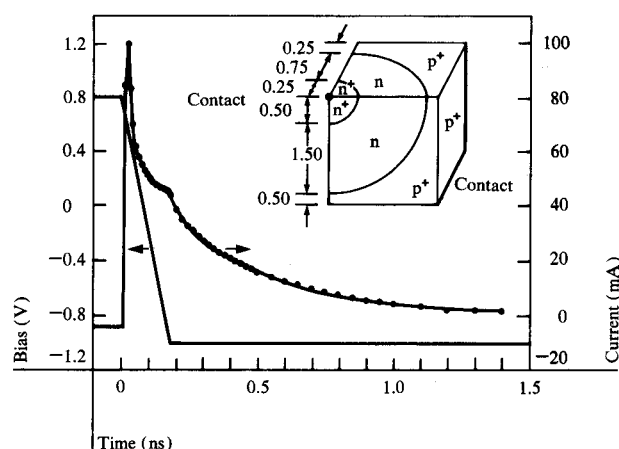


Figure 14 Reverse recovery current of an ellipsoidal junction. The bias is changed from +0.8 V to -1.0 V in 10-ps steps. The abrupt change in current at 0.0 and 0.18 ns is the displacement current which is proportional to the rate of change in bias across the junction. Dimensions in μm .

FIELDAY development efforts; G. Hachtel and R. R. O'Brien, for their guidance in formulation of the coupled solution method; and C. T. McMullen, for his contributions in the three-dimensional formulation and the modified-Newton method.

References

1. D. P. Kennedy and R. R. O'Brien, "Computer Aided Two-dimensional Analysis of the Junction Field-effect Transistor," *IBM J. Res. Develop.* **14**, 95-116 (1970).

2. M. Reiser, "A Two-dimensional FET Model for DC, AC, and Large-Signal Analysis," *IEEE Trans. Electron Devices* **ED-20**, 35-45 (1973).
3. J. W. Slotboom, "Computer-Aided Two-dimensional Analysis of Bipolar Transistors," *IEEE Trans. Electron Devices* **ED-20**, 669-679 (1973).
4. F. H. de la Moneda, "Threshold Voltage from Numerical Solution of the Two-dimensional MOS Transistor," *IEEE Trans. Circuit Theory* **CT-20**, 666-673 (1973).
5. S. P. Gaur, "Two-Dimensional Carrier Flow in a Transistor Structure Under Non-isothermal Conditions," Ph.D. Dissertation, University of Massachusetts, Amherst, 1974.
6. K. Yamaguchi, S. Asai, and H. Kodera, "Two-Dimensional Numerical Analysis of Stability Criteria of GaAs FETs," *IEEE Trans. Electron Devices* **ED-23**, 1283-1290 (1976).
7. M. H. Elsaid, S. G. Chamberlain, and L. A. K. Watt, "Computer Model and Charge Transport Studies in Short Gate Charge-Coupled Devices," *Solid-State Electron.* **20**, 61-69 (1977).
8. A. G. Fortino and J. S. Nadeau, "A Simply Implemented Dimensional Shockley-Poisson Solver with Applications to CCD," *Int. J. Numerical Methods Eng.* **11**, 251-269 (1979).
9. D. Vandorpe, J. Borel, G. Merckel, and P. Samtot, "An Accurate Two-Dimensional Numerical Analysis of the MOS Transistors," *Solid-State Electron.* **15**, 547-557 (1972).
10. H. H. Heimeier, "A Two-Dimensional Numerical Analysis of a Silicon N-P-N Transistor," *IEEE Trans. Electron Devices* **ED-20**, 708-714 (1973).
11. E. M. Buturla and P. E. Cottrell, "Two-Dimensional Finite Element Analysis of Semiconductor Steady-State Transport Equations," presented at the International Conference on Computational Methods in Nonlinear Mechanics, Austin, TX, 1974.
12. G. D. Hachtel, M. Mack, and R. R. O'Brien, "Semiconductor Device Analysis Via Finite Elements," presented at the Eighth Asilomar Conference on Circuits and Systems, Pacific Grove, CA, 1974.
13. G. W. Brown and B. W. Lindsay, "The Numerical Solution of Poisson's Equation for Two-Dimensional Semiconductor Devices," *Solid-State Electron.* **19**, 991-992 (1976).
14. E. M. Buturla and P. E. Cottrell, "Two-Dimensional Finite Element Analysis of Semiconductor Phenomena," presented at the International Conference on Numerical Methods in Electric and Magnetic Field Problems, Santa Margherita, Italy, 1976.
15. B. T. Browne and J. J. H. Miller, Eds., *Numerical Analysis of Semiconductor Devices, Proceedings of the NASECODE I Conference*, Boole Press, Dublin, Ireland, 1979.
16. A. Voshii, S. Horiguchi, and T. Sudo, "An Accurate Numerical Analysis for Very Small Semiconductor Devices," *IEEE International Solid State Circuits Conference Digest of Technical Papers*, San Francisco, Feb. 1980, pp. 80-81.
17. J. J. Barnes and R. J. Lomax, "Finite-Element Methods in Semiconductor Device Simulation," *IEEE Trans. Electron Devices* **ED-24**, 1082-1089 (1977).
18. T. Adachi, A. Yoshii, and T. Sudo, "Two-Dimensional Semiconductor Analysis Using Finite-Element Method," *IEEE Trans. Electron Devices* **ED-26**, 1026-1031 (1979).
19. G. D. Hachtel, M. H. Mack, R. R. O'Brien, and B. Speelpenning, "Semiconductor Analysis Using Finite Elements—Part I: Computational Aspects," *IBM J. Res. Develop.* **25**, 232-245 (1981, this issue); also, G. D. Hachtel, M. H. Mack, and R. R. O'Brien, "Semiconductor Analysis Using Finite Elements—Part II: IGFET and BJT Case Studies," *IBM J. Res. Develop.* **25**, 246-260 (1981, this issue).
20. E. M. Buturla, P. E. Cottrell, B. M. Grossman, M. B. Lawlor, C. T. McMullen, and K. A. Salsburg, "Three Dimensional Simulation of Semiconductor Devices," *IEEE International Solid State Circuits Conference Digest of Technical Papers*, San Francisco, Feb. 1980, pp. 76-77.
21. W. Van Roosbroeck, "Theory of Flow of Electrons and Holes in Germanium and Other Semiconductors," *Bell Syst. Tech. J.* **29**, 560-562 (1950).
22. R. J. Van Overstraeten, H. J. DeMan, and R. P. Mertens, "Transport Equations in Heavy Doped Silicon," *IEEE Trans. Electron Devices* **ED-20**, 290-298 (1973).
23. J. W. Slotboom and H. C. DeGraaff, "Measurements of Band Gap Narrowing in Si Bipolar Transistors," *Solid State Electron.* **19**, 857-862 (1976).
24. D. M. Caughey and R. E. Thomas, "Carrier Mobilities in Silicon Empirically Related to Doping and Field," *Proc. IEEE* **55**, 2192-2193 (1967).
25. C. R. Crowell and S. M. Sze, "Temperature Dependence of Avalanche Multiplication in Semiconductors," *Appl. Phys. Lett.* **9**, 242-244 (1966).
26. J. Dziewior and W. Schmid, "Auger Coefficients for Highly Doped and Excited Silicon," *Appl. Phys. Lett.* **31**, 346-348 (1977).
27. R. N. Hall, "Electron-Hole Recombination in Germanium," *Phys. Rev.* **87**, 387 (1952).
28. W. Shockley and W. T. Read, "Statistics of the Recombination of Holes and Electrons," *Phys. Rev.* **87**, 835-842 (1952).
29. W. Schottky, *Naturwissenschaften* **26**, 843 (1938).
30. C. R. Crowell and S. M. Sze, "Current Transport in Metal-Semiconductor Barriers," *Solid-State Electron.* **9**, 1035 (1966).
31. H. K. Gummel, "A Self-Consistent Iterative Scheme for One-Dimensional Steady State Transistor Calculations," *IEEE Trans. Electron Devices* **ED-11**, 445-465 (1964).
32. P. E. Cottrell and E. M. Buturla, "Two-dimensional Static and Transient Simulation of Mobile Carrier Transport in a Semiconductor," *Proceedings of the NASECODE I Conference*, Boole Press, Dublin, Ireland, 1979, pp. 31-64.
33. O. C. Zienkiewicz and Y. K. Cheung, "Finite Elements in the Solution of Field Problems," *The Engineer*, Sept. 1965, pp. 507-510.
34. J. S. Archer, "Consistent Matrix Formulations for Structural Analysis Using Influence Coefficient Techniques," presented at the First American Institute of Aeronautics and Astronautics Annual Meeting, June 1964, Paper No. 64-488.
35. P. E. Cottrell and E. M. Buturla, "Application of the Finite Element Method to Semiconductor Transport," presented at the Asilomar Conference on Circuits and Systems, Pacific Grove, CA, 1976.
36. E. M. Buturla and P. E. Cottrell, "Simulation of Semiconductor Transport Equations Using Coupled and Decoupled Solution Techniques," *Solid-State Electron.* **23**, 331-334 (1980).
37. International Business Machines Program Product, *Subroutine Library—Mathematics User's Guide*, Order No. SH12-5300-1, available through IBM branch offices.
38. P. E. Cottrell and E. M. Buturla, "Steady State Analysis of Field Effect Transistors via the Finite Element Method," *Digest of the IEEE International Electron Devices Meeting*, Washington, DC, Dec. 1975, pp. 51-54.
39. W. P. Noble and P. E. Cottrell, "Narrow Channel Effects in Insulated Gate Field Effect Transistors," *Digest of the IEEE International Electron Devices Meeting*, Washington, DC, Dec. 1976, pp. 582-586.
40. F. H. Gaensslen, "Geometry Effects of Small MOSFET Devices," *IBM J. Res. Develop.* **23**, 682-688 (1979).
41. H. Kotecha and W. P. Noble, "Interaction of IGFET Field Design with Narrow Channel Device Operation," *Digest of the IEEE International Electron Devices Meeting*, Washington, DC, Dec. 1980, pp. 724-727.
42. Richard A. Larsen, "A Silicon and Aluminum Dynamic Memory Technology," *IBM J. Res. Develop.* **24**, 268-282 (1980).
43. J. J. Barnes, D. N. Shabde, and F. B. Jenne, "The Buried-Source VMOS Dynamic RAM Device," *Digest of the IEEE International Electron Devices Meeting*, Washington, DC, Dec. 1977, pp. 272-275.

44. K. Hoffman and R. Losehand, "VMOS Technology Applied to Dynamic RAMs," *IEEE J. Solid State Circuits* SC-13, 617-622 (1978).
45. P. E. Cottrell and E. M. Buturla, "Threshold and Subthreshold Characteristics of VMOS FETs via a Two-dimensional Finite Element Model," presented at the Device Research Conference, June 1978, abstract published in *IEEE Trans. Electron Devices* ED-25, 1346 (1978).
46. D. Eardley, IBM General Technology Division laboratory, East Fishkill, NY, private communication.

Received February 2, 1981; revised March 2, 1981

The authors are located at the IBM General Technology Division laboratory, Essex Junction, Vermont 05452.